Is There an Educational Penalty for Being Suspended from School?*

Deborah A. Cobb-Clark[a]
Sonja C. Kassenboehmer[b]
Trinh Le[c]
Duncan McVicar[d]
Rong Zhang[a]

[a]**University of Melbourne, Melbourne, Australia**
[b]**Monash University, Melbourne, Australia**
[c]**Motu Economic and Public Policy Research, Wellington, New Zealand**
[d]**Queen's University Belfast, Belfast, United Kingdom**

JEL codes: I24, I28

Keywords: school suspension, school discipline, educational achievement, educational attainment, causal effect

*This paper is based on research commissioned by the Australian Government Department of Education, Employment and Workplace Relations (DEEWR) under the Social Policy Research Services Agreement (2010–2012) with the Melbourne Institute of Applied Economic and Social Research. The paper uses data from the Youth in Focus Project which is jointly funded by DEEWR, the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs and the Australian Research Council (Grant Number LP0347164) and carried out by the Australian National University. The views expressed in this paper are those of the authors alone and do not necessarily represent the views of DEEWR, the Australian Government, or the Melbourne Institute. We would like to thank Anastasia Sartbayeva and Chris Ryan for sharing their programming code, Mike Helal for help with data on academic disruptions, and DEEWR staff, the editor and two anonymous referees for helpful comments on previous drafts. Any remaining errors are those of the authors. Corresponding author, Trinh Le: email trinh.le@motu.org.nz.

Abstract

Suspension from school is a commonly-used, yet controversial, school disciplinary measure. This paper uses unique survey data to estimate the impact of suspension on the educational outcomes of those suspended. It finds that while suspension is strongly associated with educational outcomes, the relationship is unlikely to be causal, but rather likely stems from differences in the characteristics of those suspended compared to those not suspended. Moreover, there is no evidence that suspension is associated with larger educational penalties for young people from disadvantaged family backgrounds compared to those from more advantaged family backgrounds. These results hold regardless of whether self-reported suspension or mother-reported suspension is considered. The absence of a clear negative causal impact of suspension on educational outcomes suggests that suspension may continue to play a role in school discipline without harming the educational prospects of those sanctioned.

1. Introduction

Suspension, which requires a student to be absent from the classroom or school building for a specified period of time, is one of the most common, yet controversial, school disciplinary measures. In the U.S., for example, millions of students are suspended from public school each year (Kinsler, 2009). Proponents of suspension – mostly teachers and principals – argue that suspending misbehaving students can deter misbehavior (Ewing, 2000) and create an improved learning environment for other students (Public Agenda, 2004). Given that schools have a responsibility to maintain a safe and orderly learning environment for the benefit of all students, they should therefore have the autonomy to suspend misbehaving students if necessary (e.g. Public Agenda, 2004). Meanwhile, opponents – mostly parents, child advocacy and civil rights groups – argue that suspension reduces learning opportunities for affected students, impacting negatively on their academic performance and other outcomes (e.g. Losen and Gillespie, 2012). If suspension of disruptive students does improve the educational outcomes for other students, but at the cost of worse educational outcomes for those suspended – both empirical questions – then school principals face an important trade-off when making decisions about whether to use such disciplinary measures or not. Our objective is to contribute to this debate by estimating the educational penalties associated with suspension.

Suspensions are unlikely to be imposed equally across different groups of students, both because students from different groups may exhibit different behaviors on average, and because sanctions may be imposed differentially on different groups engaging in a given level of misbehavior. Jordan and Anil (2009), for example, find that low-income students are up to eight times more likely to be sent for disciplinary referrals than others. A large literature also shows that students from minority groups, especially black and Hispanic students, are disproportionately subject to school discipline (e.g. Skiba et al., 2002; Mendez and Knoff,

2003; Kinsler, 2011). Therefore, if suspension does negatively impact on an individual's educational outcomes, the use of suspension may intensify inequalities in educational outcomes.

The empirical literature concerning the impact of suspension on students' own outcomes includes contributions from several disciplines, including a handful from economics. Most are descriptive analyses which show that suspended students tend to have poorer outcomes than students who have not been suspended. If school principals are to make informed decisions regarding suspension, however, then it is critical to have evidence showing the extent to which this association is driven by a causal impact of suspension on educational outcomes as opposed to other differences between students receiving and not receiving suspensions that are themselves correlated with outcomes, i.e. selection.

This paper takes an important step in addressing this question. In particular, we examine the extent to which the negative relationship between being suspended from school and three subsequent educational outcomes – high school completion (educational attainment), achieving a university entrance score and the university entrance score achieved (educational achievement), all measured at age 20 – are reasonably interpreted as being driven by the causal impacts of suspension as opposed to selection on unobservable characteristics.

This paper also makes a second contribution by examining whether the strength of any relationships between being suspended and later educational outcomes differ by socio-economic circumstances as reflected in the family's history of welfare receipt. That is, we test whether those young people from families with a history of intensive welfare receipt are as able as their more advantaged peers to overcome the potential educational penalties associated with suspension. An educational penalty that is larger for students from disadvantaged families could lead to further inequality in educational outcomes, over and

above those resulting from differences in the prevalence of suspension across students from different backgrounds.[1]

In addressing these questions, we take advantage of unique Australian data from the Youth in Focus survey which provide detailed information about young people's educational attainment and educational achievement at age 20 along with retrospective data (including data provided by their mothers) and administrative data (from social security records) about a range of events taking place during the respondents' childhood.

We find that while suspension is strongly associated with poorer educational outcomes, the relationship is unlikely to be causal, but rather likely stems from other factors not accounted for, i.e. selection on unobservable characteristics. Moreover, there is no evidence that suspension has more negative educational consequences for young people from more disadvantaged backgrounds than for their more advantaged peers. These results hold regardless of whether self-reported suspension or mother-reported suspension is assessed.

The rest of the paper proceeds as follows. Section 2 summarizes the existing literature on the relationship between suspension and educational outcomes. The methods and data are respectively described in Sections 3 and 4. Section 5 reports and discusses the estimation results. Section 6 summarizes and concludes.

2. Background and existing literature

Suspension is widespread among U.S. school students. For example, over three million students, or about 7.4 percent of all K-12 students, were suspended at least once during the school year 2009-2010 (Losen and Gillespie, 2012). The use of suspension also appears to be on an upward trend in the U.S. For example, Wald and Losen (2003) note that the number of

[1] One of the underlying arguments for sanctioning disruptive behavior by suspension is the spillover effects that such behavior might have on other students. Estimated average peer effects in educational achievement tend to be small, but there is growing evidence that they may be heterogeneous (e.g. Sacerdote, 2011). Such differences could influence the extent to which the suspension trade-off facing school principals might vary across different groups of students.

suspensions in 2000 was about double the number in 1974. Many studies attribute this increase to high-stakes accountability policies such as 'No child left behind' and 'zero tolerance' (Imich, 1994; Leone et al., 2000; Skiba et al., 1997). There are reports that students have been suspended for relatively minor disruptive behaviors, such as having a knife ('potential weapon') in the lunch box or talking on a cell phone while at school (American Psychological Association Zero Tolerance Task Force, 2008).

There also appears to be a large race gap in the incidence of school suspension. For example, while only 7 percent of white K-12 students were suspended at least once during the school year 2009-2010, the corresponding rate for blacks was 17 percent (Losen and Gillespie, 2012). This gap has been the subject of much research and controversy.[2] The race gap in suspension came under particular attention in mid-2012 when President Obama supported a campaign to regulate schools' disciplinary actions so that members of major racial and ethnic groups are penalized at equal rates, regardless of individuals' behavior.[3]

Even though a 'zero tolerance' approach is not explicitly enforced in Australia, suspension is not uncommon. For example, in New South Wales, Australia's most populous state, 1.7 percent of all K-12 students (4.1 percent of grade 7-10 students) in 2010 were subject to a long suspension (5-20 days) (NSW Department of Education and Communities, 2010).[4] In Victoria (the second most populous state), about 2 percent of students across all grades were suspended each year during 2006-2008, with those in grades 8 to 10 more likely to be suspended (KPMG, 2010). As in the U.S., the use of suspension in Australian schools

---

[2] See Kinsler (2011) and the references cited therein.

[3] http://nation.foxnews.com/obama/2012/07/27/president-obama-backs-race-based-school-discipline-policies

[4] The same document reports that the number of short suspensions (1-4 days) was over three times the number of long suspensions but it does not report on the number of short suspended students.

attracts a great deal of attention, with recent calls in the popular media both for increased use of suspension as well as opposition to the use of suspension.[5]

The academic literature – across several disciplines including psychology, education, sociology and more recently economics – has suggested a number of channels through which suspension from school might impact negatively on later educational outcomes. These include effects on self-respect, stigma among peers, increased contact with delinquent subculture, isolation from the school setting, and the loss of instructional time incurred during the suspension (see Costenbader and Markson, 1998; Morrison et al., 2001). It seems less likely, although not impossible, that having been suspended might impact positively on later outcomes, e.g. through an improvement in behavior in response to the sanction.

Within this wider literature, a large empirical literature documents a negative relationship between suspension and later educational outcomes. For example, based on survey data for 1,300 young people, Mendez and Sanders (1981) showed that high-school graduation rates were around 40 percentage points lower for students who had been suspended relative to those who had not. Ekstrom et al. (1986) found that over 30 percent of sophomores who dropped out of school had been suspended, a rate three times that of their peers who stayed in school. A weaker relationship is observed by Mendez (2003) who uses data tracking around 8,000 school pupils over a 13-year period to show that suspension is (moderately) negatively correlated with on-time graduation. None of these studies, however, cast light on the extent to which these associations can be interpreted as demonstrating causal relationships between suspension and educational attainment.[6]

---

[5] See, for example, http://www.thechronicle.com.au/story/2012/07/19/school-incidents-spark-call-for-more-discipline and http://www.theaustralian.com.au/national-affairs/education/suspend-judgment-keep-kids-at-school/story-fn59nlz9-1226398136400, respectively.

[6] At best this literature examines the association between suspension and achievement matching on or otherwise controlling for a small number of observable characteristics (e.g. gender, race) either at the individual or aggregate level (e.g. Rausch and Skiba, 2005; Arcia, 2006).

In contrast, Kinsler (2013) attempts to identify the causal impact of suspension on educational achievement, as measured by end of grade tests, for middle-school students in grades 6-8. He exploits a rich data set on test scores, behavioral infractions that lead to suspension, and suspension durations across three school districts in North Carolina to estimate a model for the joint determination of behavior, suspension and achievement. To mitigate the effect of selection on unobservable characteristics in this model, he assumes students are drawn from a finite mixture distribution of types (i.e. that there are essentially two types of students – 'bad seeds' and 'good seeds') and controls for student type in his model. Using this approach, Kinsler finds a large apparent negative effect of each day of suspension on own achievement which disappears when he controls for student type. He concludes that suspension does not have a negative causal impact on own achievement.[7] Our approach to the selection problem differs from that of Kinsler (2013). We estimate a series of models, increasing in controls, to gauge the stability of our results to omitted variables. We then examine the sensitivity of our estimates to any systematic error in self-reports of suspension using information from mothers' reports of suspension. Finally, we undertake a sensitivity test suggested by Altonji et al. (2005) to assess the potential for selectivity bias to be driving our results. To our knowledge, no other study explicitly focuses on the extent to which the observed relationships between suspension and later educational outcomes can reasonably be interpreted as causal.[8]

---

[7] In contrast, he argues that suspension does have positive impacts both as a deterrent to misbehavior and on the achievement of other class members via peer effects. The implication is that school principals do not face a trade-off between outcomes for the individual and outcomes for the rest of the class in decisions regarding the use of suspension.

[8] In a related study Karakus et al. (2012) estimate a recursive bivariate probit model to control for direct effects of behavioral problems on employment as well as the indirect effects through endogenous high school graduation. They show that middle-school behavior problems (although not explicitly suspension) impact negatively on high-school graduation but, conditional on graduation, have no significant impact on adult employment.

3. Data and descriptive statistics

3.1. Youth in Focus Survey

This study draws on data from the Youth in Focus (YIF) Survey, which asks questions about family background, living arrangements, education, work, relationships, income, health, leisure time, and aspirations and attitudes of young people in Australia. Individuals born between October 1987 and March 1988 who appear in the Australian social security administrative database (Centrelink) between 1993 and 2005 were randomly selected and invited to participate in the survey. Young people have a Centrelink record if they receive any government payment (e.g. Youth Allowance) in their own right. They are more likely, however, to have a Centrelink record because someone in their family (usually a parent) received a payment between 1993 and 2005 which depended in part on his/her relationship to the youth. Because Australian social security benefits are nearly universal for families with children, over 98 percent of young people born between October 1987 and March 1988 in the overall Australian population appear in the Centrelink sampling frame (Breunig et al., 2009).

Respondents were first interviewed in late 2006 (wave 1) when they were around 18 years of age and then in late 2008 (wave 2) around age 20. Wave 2 respondents include both continuing respondents (those who participated in wave 1) and new entrants. In this paper, we examine educational outcomes measured at age 20 (wave 2) in order to allow time for individuals born later in the cohort, entering school late, or repeating a grade to have enough time to have completed high school.

A particular strength of the YIF data is that a responsible adult – in 96.5 percent of cases the biological mother – for each respondent was also invited to answer the parent questionnaire. Wave 2 included 3,623 young respondents, 1,879 of whom had matched parental records. In order to get consistent information on parents, we restrict the sample to

those young people whose biological mothers also answered the parent questionnaire. This excludes 50 cases. Also excluded are cases where the parent had not lived with the young respondent in the past five years (8 cases), and where the young respondent was still attending school at wave 2 (5 cases). As a result, the maximum sample size is 3,560 for analyses that do not use mother's information and 1,819 for analyses that do.[9]

3.2. Outcomes

We consider three educational outcomes: high school completion, achieving a university entrance score, and the entrance score itself. Australian students who complete secondary school (high school) and who meet certain minimum coursework requirements (e.g. with respect to minimum credit hours, English language requirements, etc.) are assigned a percentile ranking (from 1 – 100) based on their academic performance in grades 11 and 12. In some states and territories, this ranking is derived solely from a state-wide exam. In others, the final results of specific subjects are used in combination with standardized tests.[10] Students wishing to attend university register their preferences (in order) for the various programs at particular universities. University offers are then made centrally on the basis of students' entrance scores (Marks et al., 2001).

---

[9] Non-response to particular items in the questionnaire means sample size is usually lower than these two figures. Sample size is further reduced where we analyse outcomes whose occurrence depends on the occurrence of another outcome.

[10] University entrance scores are known by different names in different States and Territories in Australia. Queensland uses a different system called the Overall Position (OP). The OP score ranges between 1 and 25, where 1 is the highest and 25 is the lowest possible score. In all other states the score ranges from 1 to 100 (highest). We transform the OP score to match the other scores using the conversion factors that university administrators use when comparing Queensland school leavers with those from other states for the purpose of university admission. Scores under 30 are reported as being 30 to the student, so reported scores of under 30 must be data errors and we recode these to missing here.

3.3. Key control variables

We have a number of observed measures of socio-economic disadvantage available. But as in Cobb-Clark et al. (2012), here we are particularly interested in a proxy for disadvantage based on family welfare-receipt history, which categorizes families as follows:

- those with no history of welfare receipt;
- those that received less than six years of welfare after 1998 when the respondent was older than 10 (late moderate welfare receipt);
- those that received less than six years of welfare, some of which occurred before 1998 when the respondent was younger than 10 (early moderate welfare receipt); and
- those that received welfare for more than six years while the respondent was growing up (intense welfare receipt).

We also control for respondents' other schooling experiences (number of schools attended and whether the respondent had repeated a year), respondents' demographic characteristics, parental characteristics, and family characteristics when the respondent was 14 years of age, including whether the respondent lived with both parents, and whether the respondent's mother was employed.

3.4. Descriptive statistics

Table 1 contains sample means for the educational outcomes, schooling experiences, and control variables. Over 79 percent of respondents completed high school. Among those completing high school, 81 percent met the curriculum requirements to be awarded a university entrance score, with a mean score of 75.[11] Over a quarter of respondents grew up in families with a history of intensive welfare receipt, while under a quarter (8 percent) live in

[11]The mean university entrance score is inflated, as scores below 30 are reported as 30.

families with moderate early (late) receipt. The results in Table 1 also suggest that individuals with matched parental records have somewhat different observable characteristics to the total sample.[12]

Suspension is relatively common, with almost 19 percent of all respondents reporting having been suspended at least once during their schooling. Individuals who have been suspended from school compare unfavorably to others on all three educational outcomes. Only 54 percent of those who have been suspended from school go on to complete high school in comparison to 85 percent of non-suspended students. Of those completing high school, 83 percent of those not suspended achieve a university entrance score compared to 67 percent of those suspended from school. There is also an 8 percentile point achievement gap between suspended and non-suspended students in terms of the university entrance scores achieved.

Suspension is also associated with other observable factors. Among those respondents who have been suspended from school, 43 percent are from families with intensive welfare-receipt history, whereas the corresponding proportion is only 24 percent among those who have not been suspended. Equivalently, 30 percent of respondents with intensive welfare-receipt history have been suspended from school, whereas only 15 percent of those from families with moderate or no welfare receipt have been suspended. In other words, in Australia as in the U.S., suspension is more prevalent among young people from disadvantaged family backgrounds.

---

[12] For example, the proportion reporting a suspension is higher in the unmatched sample than in the matched sample. We cannot rule out that there is also selection into the matched sample on relevant unobservables, such as frequency or severity of suspensions among those reporting suspension. If it is the more serious suspension cases in the unmatched sample and/or in the original population that are dropping out of the matched sample then this could potentially limit the extent to which our conclusions apply to such cases.

3.5. Comparison of youth's and mother's reports of suspensions

A unique feature of the YIF data is that information about young people's schooling experiences is collected from both the mother and the young person. We therefore have an alternative measure of suspension available, i.e. as reported by mothers rather than the young people themselves. There is considerable inconsistency between the two measures, both of which may measure suspension with error. Such measurement error could be random (forgetfulness) or systematic (over-reporting or under-reporting), with different implications for the interpretation of our estimates. Mothers report that 11.2 percent of responding (and matched) young people have been suspended from school, while the corresponding rate reported by young people themselves is 19.4 percent. Among cases in which both the young person's and the mother's reports are available, both reports agree in 91 percent of the time. Where reports disagree, it is more common that the young person reports a suspension that the mother does not rather than the converse. So mothers under-report suspensions relative to the young people themselves (either more under-reporting or less over-reporting), but otherwise reports appear consistent. Our conjecture is that this primarily reflects shorter suspensions for more minor incidents of misbehavior that the mother either did not know about at the time or has forgotten, in which case suspensions reported by mothers can be thought of as more 'serious' on average than suspensions reported by the young people. Other interpretations of this inconsistency are also possible, however, such as 'bragging' by young people or embarrassment on the part of mothers of suspended children. In what follows we examine sensitivity to using either mother or young person reports of suspension.

4. Estimation strategy

Our starting point is the following reduced-form model:

$$Y_i = \alpha \quad + \beta_S S + \boldsymbol{\beta_X X} + \varepsilon_i \tag{1}$$

where i indexes individuals, S is a binary indicator of whether or not the individual was ever suspended from school, and $\boldsymbol{X}$ is a vector of controls for own, family and parental characteristics that are related to educational outcomes Y; α, β_S and $\boldsymbol{\beta_X}$ are parameters to be estimated. The parameter of interest is β_S which captures the gap in educational outcomes for those who report having been suspended and those that do not, holding other observable characteristics constant. Note that we do not observe reported misbehavior separately from reported suspension. As a result, without further assumptions, the estimated β_S resulting from equation (1) cannot be interpreted literally as the effect of having been suspended conditional on *a given level of misbehavior*. Equation (1) is estimated using a probit model (and the average marginal effects are derived) for our binary outcomes – high school completion and achieving a university entrance score. An ordinary least squares (OLS) model is used for our continuous measure of university entrance scores.[13]

It is important to note that the estimated determinants of the probability of achieving a university entrance score and the entrance score itself are only representative of the sample of high school completers and high school completers who achieve an entrance score, respectively, and cannot be generalized to the whole population of 20-year olds, even in the YIF study, as these subsamples are unlikely to be randomly drawn from the larger population. As a result, it is likely that our models underestimate the negative effect of suspension (for the wider population) since it is reasonable to expect that those who fail to complete high school (or fail to achieve an entrance score upon completion) have unobserved characteristics that are positively correlated with both dropping out and suspension.[14]

---

[13] As a sensitivity check, we also estimated a tobit model in order to account for the fact that university entrance scores are censored at 30 and 100. The results are virtually identical to the OLS estimates as data censoring affects very few respondents.

[14] We do not apply a Heckman selection model as it is not feasible to find a valid exclusion restriction that influences one educational outcome (such as high school completion) but not another closely related educational outcome (achieving a university entrance score given high school completion).

In order to interpret the estimated $\widehat{\beta_s}$ resulting from equation (1) as the causal effect of having been suspended on educational outcomes, we must assume that being suspended is exogenous with respect to educational attainment and achievement. A causal estimate is necessary to inform school principals about the trade-offs between the individual costs and the collective benefits of suspension. There are many reasons, however, to suspect that young people who report being suspended differ in unobserved ways to those who do not. For example, on average those young people who report having been suspended may be likely to have misbehaved at school more often (or more intensively) than their peers who do not report having been suspended. Young people who do and do not report being suspended may also differ in terms of unobserved ability, self-control, risk preferences, etc. Neither can we rule out that having been suspended may be reported with systematic error, whether we use self-reports or mothers' reports. Consequently, in a robustness analysis, we investigate how likely it is that we overestimate the true negative effect of suspension when estimating equation (1) because unobserved individual (e.g. behavior, ability) or family factors (e.g. parental support) are correlated with both reported suspension and our educational outcome variables. In other words, we investigate how much of our $\widehat{\beta_s}$ from equation (1) is likely to be driven by selection on unobservable characteristics.

In the absence of panel data or a credible instrumental variable we adopt three alternative strategies for dealing with the potential endogeneity problem. First, exploiting the richness of the YIF data, we estimate different versions of (1) increasing in controls in order to assess how stable our results are to the inclusion of observable individual, parental, and family background characteristics. Second, to examine the sensitivity of our estimates to potential systematic error in self-reports of suspension, we re-estimate our prefered model using mothers' reports of suspension in place of self-reports. Third, we conduct a sensitivity analysis proposed by Altonji et al. (2005) which allows us to gauge the potential role of any

remaining selection bias in driving our estimates. In effect, we investigate how strong the role of selection on unobservables (relative to selection on observables) would have to be in order for us to completely attribute our estimates to the effect of selection into reported suspension. The advantage of this approach is that no exclusion restrictions are needed. The disadvantage is that we learn only whether it is *reasonable* to expect some of the estimated effect to be unaffected by selection on unobservable characteristics.

5. Estimation results

5.1. Educational attainment: high school completion

Average marginal effects of reporting suspension on the probability of completing high school are presented in Table 2. (The full sets of average marginal effects, including for the controls, are presented in Appendix Table 1.) The first column presents the results from a basic model where only (self-reported) own schooling experiences (ever repeated a year, and number of schools attended) and characteristics are included as explanatory variables. The second specification (column 2a) adds further controls for family characteristics when the respondent was 14 years of age and family welfare history. As a robustness check we re-estimate this same model on the smaller sample of respondents with matched parental records only (column 2b). In the third specification (column 3), parental characteristics (mostly taken from the parent questionnaire) are added. The final specification is identical to the third, except that mother's reports of youth's schooling experiences are used in place of youth's reports.

The first specification (column 1) shows that respondents' self-reported suspension is highly correlated with high school completion, even after controlling for own characteristics. Those who report having been suspended from school are 26 percentage points less likely to complete high school. This marginal effect is larger in absolute magnitude than the marginal

effect for any of the controls (in all versions of the model); the next biggest marginal effect is for grade repetition which is associated with a 17 percentage point lower probability of completing high school. Despite its magnitude, this result points to a smaller attainment gap in Australian than that suggested for the U.S. by, for example, Mendez and Sanders (1981, 40 percentage points). The estimated effects of all of the demographic controls are statistically significant and take expected signs. For example, one of the largest associations is that the probability of high school completion is 17 percentage points lower for Indigenous Australians compared to others.

The controls for family characteristics measured when the respondent was 14 years of age and family welfare history (column 2a) all are statistically significant. Again, the marginal effects of these additional controls take expected signs. For example, living with both parents at the age of 14 is associated with a 6 percentage point higher probability of high school completion, while those young people whose mothers were employed when they were 14 are 2.4 percentage points more likely to complete high school. Compared to those who have no family history of welfare-receipt, those with a history of intensive receipt are 8 percentage points less likely to complete high school. The corresponding effects for 'moderate (early) receipt' and 'moderate (late) receipt' are 5.2 and 5.9 percentage points, respectively. Including these additional controls reduces the estimated marginal effect of reported suspension (in absolute terms) from 26 percentage points to 18 percentage points. There are also some changes to the estimated marginal effects of the controls already included in specification 1.

Column 2b of Table 2 shows the marginal effect of suspension for the same model as in column 2a but estimated on the smaller sample of matched parent-child records. The results from this specification (column 2b) are broadly similar to those from the same specification estimated on the larger sample (column 2a), and the estimated marginal effect of suspension

changes very little. The implication is that the differences between the larger and smaller samples are largely orthogonal to educational attainment, so that sample selection bias resulting from using the smaller matched sample is unlikely to be a major issue.[15]

Adding parental characteristics to the model estimated on the smaller matched sample (column 3) reveals that many are significantly related to high school completion in the expected direction. For example, those young people whose fathers have at least some tertiary education are 6.3 percentage points more likely to complete high school. Once again the estimated marginal effect of suspension changes only very slightly when these additional controls are included. In contrast, the estimated effects of some of the existing controls fall in magnitude and in some cases become statistically insignificant. In other words, these parental characteristics appear to have absorbed some of the effects attributed to youths' own characteristics, schooling experiences and family history in the previous specification.

Since adding these parental characteristics reduces the estimation sample size without noticeable improvement in the model's goodness of fit, the remainder of this study uses the second specification (column 2a) as the benchmark specification.

The final variant of the model in Table 2 (column 4) replaces self-reports of schooling experiences (suspension, repetition and the number of school attended) with mother's reports and adds parental characteristics. Comparing columns 2a and 4, the effect of suspension is slightly higher when using mothers' reports instead of youths' own reports, but the difference is small. While self-reported suspension is associated with a reduction of 18 percentage points in the probability of completing high school, the corresponding effect for mother-reported suspension is 19 percentage points. The implication is that bias due to measurement

[15] Note, however, that standard errors in column 2b are higher than in column 2a, indicating that the smaller sample size results in less precise estimates.

error in reported suspensions is not a major issue here.[16] The effects of other controls change little.

5.2. Educational achievement: university entrance scores

Suspension is also significantly related to meeting the requirements to obtain a university entrance score as well as the entrance score achieved (Table 2, with full results for our prefered specifications report in Appendix Table 2). Conditional on completing high school, (self-reported) suspension from school is associated with a 9.8 percentage point decrease in the probability of obtaining an entrance score in our prefered specification (column 4 in Table 2). Conditional on obtaining an entrance score, suspension is also associated with a 6.8 point lower entrance score (column 4 in Table 2). The effect of self-reported suspension on both outcomes is highly robust to the inclusion of extra controls and to replacing or instrumenting self-reported suspension with mother-reported suspension. As for educational attainment, the marginal effect of suspension is larger than that for nearly any other variable for both outcomes. The exception is the effect of intensive family welfare receipt on the probability of achieving an entrance score, although only when self-reported suspension is used.

In fact, relatively few of the controls in the prefered models for achievement are statistically significant, in part because point estimates are small but also in part because the smaller sample sizes once we condition on high school completion and achieving a university entrance score result in a lack of precision (see Appendix Table 2). Family welfare history is a strong determinant of the probability of achieving a university entrance score. Compared to respondents whose family has no history of welfare receipt, those with intensive family

[16] We also estimate an IV version of the model where self-reported suspension is instrumented by mother-reported suspension. Although arguably not an ideal instrument, the IV results are very similar to those in Table 2 with suspension associated with an estimated 19 percentage point reduction in the probability of completion. We thank an anonymous referee for this suggestion.

welfare receipt are 8.7 percentage points, and those with moderate (early) family welfare receipt 6.1 percentage points, less likely to achieve an entrance score. However, family welfare receipt has no significant effects on the entrance score achieved.

Recall that because these achievement models are estimated on selective samples – the probability of achieving an entrance score is estimated on those who have completed high school, and the score achieved is estimated on those who achieved one – these estimates are likely to serve as 'conservative' estimates of the conditional associations we might expect for the total population.

5.3 Selection on unobservable characteristics

The estimates presented in column 5 of Appendix Table 2 indicate that suspension is associated with many observed characteristics. Specifically, young people who have repeated a year are more likely to have been suspended. Also strongly positively associated with suspension are the number of schools attended, intensive welfare receipt and maternal smoking. By contrast, girls and those whose fathers have at least Year 12 education are less likely to have been suspended. These findings are similar to Kinsler's (2013) findings that students who have repeated a grade, who are male or who come from less educated families are more likely to commit infractions. Interestingly, whereas Kinsler (2013) finds that black students are more likely to commit infractions, this study finds that indigenous students are no more likely to be suspended, while those from a non-English speaking background (NESB) are much less likely to have been subject to this sanction.

It is possible that suspension is also associated with unobserved characteristics which might be correlated with educational outcomes. In this section we assess the extent to which the relationships between suspension and subsequent educational outcomes discussed above might be driven entirely by unobserved differences between those who have been suspended

and those who have not, i.e. selection bias. Specifically, we exploit the fact that we control for a rich set of observed characteristics to adopt the approach suggested by Altonji et al. (2005).[17]

Table 3 presents results from constrained simultaneous equations models, which take the form of bivariate probit models for binary outcomes (high school completion and obtaining a university entrance score) and systems combining a probit regression and a linear regression for continuous outcomes (the entrance score achieved). In each case the first equation is our preferred specification for equation (1) and the second equation regresses the (self-reported or mother-reported) suspension dummy on the set of observable characteristics and an error term that captures the effect of unobservable characteristics on the probability of reporting suspension. The correlation coefficient (ρ) of the error terms in each system of regressions is constrained to be constant at a series of values ranging between -0.1 and -0.5. These values of ρ reflect those in Altonji et al. (2005), although we impose negative signs because we assume that unobserved characteristics that have a positive effect on suspension (e.g. propensity for misbehavior) are likely to have a negative effect on subsequent educational outcomes. In effect, we assume the selection bias is negative. Assuming a zero ρ is equivalent to estimating the reduced form given by equation (1), i.e. treating suspension as exogenous (as in Table 2). Higher values (in absolute terms) of ρ impose a higher correlation in those unobserved factors that shape both educational outcomes and suspension, i.e. more selection bias.

The results in Table 3 show that the estimated effects of suspension on all three education measures are highly sensitive to the degree of correlation imposed on the

[17] Because two of our three models are non-linear, and because conclusions in the remaining linear model (for university entrance score) are highly sensitive to the value assumed for R_{max}, we do not implement the related Oster (2013) approach to infer selection bias by examining coefficient stability and movements in R^2 as additional controls are added to the model.

unobserved determinants of suspension and educational attainment or achievement. For example, at $\rho = -0.1$ the negative association between suspension and achieving a university entrance score halves in magnitude and becomes statistically insignificant at the 95% level. Across all three outcomes, using either both mother- and self-reported suspension, $\rho = -0.3$ is sufficient to lead the negative relationship between suspension and later outcomes to either lose statistical significance or change sign. The most persistent effect is found for high school completion, where a marginally statistically significant negative association between suspension and completion disappears only when we impose a relatively high degree of selection ($\rho = -0.4$). Taken together, these results suggest that the effect of suspension on each of our educational outcomes is potentially completely due to the effects of selection bias.

To investigate this further, we next follow Altonji et al. (2005) in calculating the degree of selection on unobservable relative to observable characteristics that would be required to generate a selectivity bias so big that it could potentially explain all of the estimated relationship between suspension and subsequent educational outcomes. These authors argue that, given a reasonably rich set of observables, a ratio of estimate to bias of less than 1 is likely to indicate an estimated association between two variables that could plausibly be entirely explained by selection bias. In contrast, a ratio of estimate to bias of greater than 1 is likely to indicate an estimated relationship that is not completely explained by selection bias.

The results presented in Table 4 show that for all three outcomes, and for both measures of suspension, the ratio of estimate to bias is between 0.29 and 0.48. In other words, it only takes selection on unobserved characteristics to be 0.29-0.48 times as large as selection on the observed characteristics to explain all of the estimated effects of suspension by selection bias. Even the relationship between suspension and high school completion falls easily at this hurdle. Thus, under the Altonji et al. (2005) assumptions, any effect of

suspension on subsequent educational attainment and achievement conditional on own, family, and parental characteristics and family welfare history (see Table 2) appears unlikely to be causal, but more likely stems from unobservable differences between those suspended and those not suspended. Despite the differences in the approach taken, the institutional context, and the different nature of the sample, this conclusion is very much in line with that of Kinsler (2013), who similarly concludes that suspension has no causal impact on educational outcomes.

5.4 Differential impact of suspension by socio-economic circumstances

One of our objectives is to assess whether the effect of suspension on young people's educational outcomes varies by the socio-economic circumstances of their families. Our conjecture is that the educational penalty associated with suspension may be larger for young people living in disadvantaged households. For example, it has been argued that children living in households with a low socio-economic status tend to have less parental supervision than children in households with a high socio-economic status (e.g. Zick and Allen, 1996). Thus, any negative impacts of suspension might be exacerbated for children from relatively disadvantaged families.

As discussed in Section 3.3, our proxy for socio-economic disadvantage is based on families' welfare histories. To simplify, in this analysis we distinguish between only two categories of families: those who received intensive welfare support (i.e. more than six years) and those who did not. Thus, socio-economic disadvantage is proxied by intensive welfare receipt and socio-economic advantage is proxied by either no or moderate welfare receipt. We then estimate an extension of (1) that includes an interaction between our binary indicator of disadvantage and the suspension dummy. A statistically insignificant interaction implies that the relationship between educational outcomes and suspension does not vary by this measure of socio-economic disadvantage. Note that care must be taken in generating the

correct marginal effects for non-linear models, as pointed out by Norton et al. (2004). Here we calculate marginal effects for the interaction terms using the method described in Ai and Norton (2003) and Karaca-Mandic et al. (2012).

The results (panel 1a of Table 5) indicate that suspension is negatively related to educational outcomes for both advantaged and disadvantaged young people. At the same time, the interaction effect is statistically insignificant – and small in magnitude – regardless of the educational outcome we consider. Thus while suspension is negatively related to the educational outcomes of each group of young people, the magnitude of these educational penalties are not statistically different from one another. To test the robustness of these results, we redefine "disadvantage" to be having any history of welfare receipt. This definition doubles the number of individuals to whom we assign a disadvantaged background. However, the statistical patterns remain the same: the interaction term is small and insignificant for all outcomes (panel 1b of Table 5).[18] When mothers' reports of suspension are used, the point estimates remain largely the same, while the standard errors increase (as mothers report fewer suspensions than do young people themselves). As a result, the estimated effects are even less significant. Taken together there is little evidence in support of the conjecture that the effect of suspension on later outcomes is more negative for disadvantaged young people than for their more advantaged peers.[19]

6. Conclusions

The widespread use of suspension from school as a disciplinary measure has become the subject of a highly-charged debate in the education sector in the United States and beyond.

[18] In a further robustness check we redefine "disadvantage" to be having intensive welfare receipt versus no welfare receipt, dropping everyone else from the sample. Once again, the interaction term is small and insignificant for all outcomes.

[19] We also use occupational ranking (the ANU4 scale) of the mother as an alternative proxy for socio-economic disadvantage. The ANU4 scale is a continuous measure developed at the Australian National University (for more details see Jones and McMillan, 2001). Here the ANU4 scale is calculated from the current or most recent occupation. Our conclusion remains the same. Results are available upon request.

Quantitative research on the effects of suspension on later educational outcomes for those students suspended can help to inform that debate, and more specifically, can help education officials weigh up the longer-run educational costs imposed on suspended students with the educational benefits resulting from either the deterrent effect of suspension on disruptive behavior or the temporary removal of disruptive pupils from the classroom. Unfortunately, if this is our aim it is not enough to compare outcomes for those suspended with those not suspended – and much of the empirical literature on this issue, across multiple disciplines, goes no further than this – because differences in outcomes will reflect differences in the characteristics of those suspended and those not suspended in addition to any causal impact of suspension on outcomes. Kinsler (2013) is unique in attempting to identify the causal impact of having been suspended on educational achievement. Here we take a different methodological approach to Kinsler (2013), but we reach broadly the same conclusions.

Specifically, we find that having been suspended from school is strongly related to the chances of completing high school, the chances of obtaining a university entrance score, and to the university entrance score achieved, even when a host of other observed schooling experiences, own, family and parental characteristics, and family welfare-receipt history are controlled for. This is consistent with the majority of the existing literature which finds a strong relationship between suspension and later outcomes, especially high school completion. While students from disadvantaged family backgrounds are more likely to report having been suspended than are their more advantaged peers, the educational penalties associated with suspension are statistically equivalent for both groups. Finally, we exploit the unique strength of our data to demonstrate that these relationships are robust to using either self-reported or mother-reported measures of suspension.

However, using a method first proposed by Altonji et al. (2005), we also find that the strong links between suspension and educational attainment and achievement are highly

likely to be driven by selection bias, i.e. by differences in unobservable characteristics between those suspended and those not suspended. In other words, despite these strong statistical relationships, the evidence suggests that being suspended is unlikely, on average, to have a significant negative causal impact on later educational outcomes. This suggests that it is reasonable for suspension to continue to play a role in school discipline as it does not appear to harm the educational outcomes of those sanctioned. To put it another way, although we do not examine the other side of the potential trade-off facing school principals in reaching decisions regarding suspension – the possible benefits of suspension – there is certainly no evidence here that suggests school principals have got the balance wrong on the side of excessive use of suspension. Moreover, despite the fact that children from disadvantaged backgrounds are disproportionately suspended, this in itself appears unlikely to contribute to socio-economic disparities in educational attainment and achievement.

References

Ai, C. and Norton, E. (2003). Interaction terms in logit and probit models. *Economics Letters*, 80(1):123–129.

Altonji, J. G., Elder, T. E. and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, 113(1):151–184.

American Psychological Association Zero Tolerance Task Force (2008). Are zero tolerance policies effective in the schools?: An evidentiary review and recommendations. *American Psychologist*, 63(9):852–862.

Arcia, E. (2006). Achievement and enrollment status of suspended students: Outcomes in a large, multicultural school district. *Education and Urban Society*, 38: 359-369.

Breunig, R., Cobb-Clark, D., Gorgens, T., Ryan, C. and Sartbayeva, A. (2009). User's guide to the Youth in Focus data, version 2.0. Youth in Focus Project Discussion Paper Series 8, Research School of Economics, Australian National University.

Cobb-Clark, D., Ryan, C. and Sartbayeva, A. (2012). Taking chances: The effect of growing up on welfare on the risky behaviour of young people. *Scandinavian Journal of Economics*, 114(30):729–755.

Costenbader, V. and Markson, S. (1998). School suspension: A study with secondary school students. *Journal of School Psychology*, 36(1):59–82.

Ekstrom, R., Goertz, M., Pollack, J. and Rock, D. (1986). Who drops out of high school and why? Findings from a national study. *The Teachers College Record*, 87(3):356–373.

Ewing, C. P. (2000). Sensible zero tolerance protects students. Harvard Education Letter.

Imich, A. (1994). Exclusions from school: Current trends and issues. *Educational Research*, 36(1):3–11.

Jones, F. and McMillan, J. (2001). Scoring occupational categories for social research: A review of current practice, with Australian examples. *Work, Employment & Society*, 15(3):539–563.

Jordan, J. and Anil, B. (2009). Race, gender, school discipline, and human capital effects. *Journal of Agricultural and Applied Economics*, 41(2):419–429.

Karaca-Mandic, P., Norton, E. and Dowd, B. (2012). Interaction terms in nonlinear models. *Health Services Research*, 47(1):255–275.

Karakus, M. C., Salkever, D. S., Slade, E. P., Ialongo, N. and Stuart, E. (2012). Implications of middle school behaviour problems for high school graduation and employment outcomes of young adults: Estimation of a recursive model. *Education Economics*, 20(1):33–52.

Kinsler, J. (2009). School policy and student outcomes in equilibrium: Determining the price of delinquency. University of Rochester.

Kinsler, J. (2011). Understanding the black-white school discipline gap. *Economics of Education Review*, 30(6):1370–1383.

Kinsler, J. (2013). School discipline: A source or salve for the racial achievement gap. *International Economic Review*, 54(1), 355–383.

KPMG (2010). Re-engaging our kids a framework for education provision to children and young people at risk of disengaging or disengaged from school. Prepared for Department of Education and Early Childhood Development.

Leone, P., Mayer, M., Malmgren, K. and Meisel, S. (2000). School violence and disruption: Rhetoric, reality, and reasonable balance. *Focus on Exceptional Children*, 33(1):1–20.

Losen, D. and Gillespie, J. (2012). Opportunities suspended: The disparate impact of disciplinary exclusion from school. The Civil Rights Project, UCLA.

Marks, G. N., McMillan, J. and Hillman, K. (2001). Tertiary entrance performance: The role of student background and school factors. *Research Report Number 22,* Australian Council for Education Research.

Mendez, L. and Knoff, H. (2003). Who gets suspended from school and why: A demographic analysis of schools and disciplinary infractions in a large school district. *Education and Treatment of Children*, 26(1):30–51.

Mendez, L. M. (2003). Predictors of suspension and negative school outcomes: A longitudinal investigation. *New Directions for Youth Development*, 2003(99):17–33.

Mendez, R. and Sanders, S. (1981). An examination of in-school suspension: Panacea or Pandora's box? *NASSP Bulletin*, 65(441):65–69.

Morrison, G., Anthony, S., Storino, M., Cheng, J., Furlong, M. and Morrison, R. (2001). School expulsion as a process and an event: Before and after effects on children at risk for school discipline. *New Directions for Mental Health Services*, 2001(92):45–71.

Norton, E., Wang, H. and Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. *Stata Journal*, 4:154–167.

NSW Department of Education and Communities (2010). Long suspensions and expulsions 2010.

Oster, E. (2013). Unobservable selection and coefficient stability: Theory and validation. NBER Working Paper No. 19054, Cambridge, MA.

Public Agenda (2004). Teaching interrupted: Do discipline policies in today's public schools foster the common good? Public Agenda Foundation and Common Good Foundation, New York, NY.

Rausch, M.K. and Skiba, R.J. (2005). The academic cost of discipline: the relationship between suspension/expulsion and school achievement. Paper Presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada, April, 2005.

Sacerdote, B. (2011). 'Peer effects in education: how might they work, how big are they and how much do we know thus far?' In Hanushek, E., Machin, S. and Weossman. L. (eds.) *Handbook of the Economics of Education Volume 3,* Amsterdam: Elsevier.

Skiba, R., Peterson, R. and Williams, T. (1997). Office referrals and suspension: Disciplinary intervention in middle schools. *Education and Treatment of Children*, 20(3):295–315.

Skiba, R., Michael, R., Nardo, A. and Peterson, R. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The Urban Review*, 34(4):317–342.

Wald, J. and Losen, D. (2003). Defining and redirecting a school-to-prison pipeline. *New Directions for Youth Development*, 2003(99):9–15.

Zick, C. D. and Allen, C. R. (1996). The impact of parents' marital status on the time adolescents spend in productive activities. *Family Relations*, 45(1):65–71.

Table 1: Means of key regression variables

	Sample		Suspended from school (total sample)	
	Total sample	With matched parental records only	No	Yes
Outcomes				
Completed high school	0.792	0.816***	0.852	0.542***
Achieved a university entrance score	0.812	0.831***	0.833	0.665***
University entrance score (scale: 0-100)	75.008	75.624*	75.761	67.949***
Schooling experiences (own report)				
Suspended from school	0.194	0.175***	0	1
Repeated a year	0.105	0.101	0.088	0.177***
Number of schools	2.948	2.814***	2.847	3.369***
Own characteristics				
Female	0.490	0.494	0.532	0.314***
Indigenous Australian	0.033	0.029	0.026	0.057***
Born in a non-English speaking (NES) country	0.077	0.038***	0.085	0.044***
Metropolitan residence	0.666	0.627***	0.676	0.628**
Family characteristics at age 14				
Lived with both parents at 14	0.738	0.782***	0.775	0.588***
Mother employed at 14	0.703	0.736***	0.710	0.670**
Mother had Year 12 when youth 14	0.500	0.527***	0.524	0.400***
Mother had Year 12 when youth 14: missing	0.074	0.058***	0.062	0.120***
Father had Year 12 when youth 14	0.491	0.498	0.515	0.383***
Father had Year 12 when youth 14: missing	0.070	0.056***	0.062	0.106***
Family welfare history				
Moderate (late) receipt	0.081	0.076	0.084	0.070
Moderate (early) receipt	0.234	0.237	0.236	0.224
Intensive receipt	0.275	0.229***	0.237	0.433***
Parental characteristics				
Age of mother	47.039	47.039	47.156	46.453**
Number of children of mother	2.951	2.951	2.937	3.017
Mother is a smoker	0.188	0.188	0.160	0.319***
At least one parent born in a NES country	0.241	0.183***	0.250	0.205**
At least one parent born in a NES country: missing	0.010	0.009	0.007	0.023***
Mother's education: Year 12	0.086	0.086	0.092	0.059*
Mother's education: above Year 12	0.654	0.654	0.648	0.678
Father's education: Year 12	0.149	0.149	0.156	0.114*
Father's education: above Year 12	0.461	0.461	0.492	0.313***
Number of observations	3,560	1,819	2,839	715

Source: Australian Youth in Focus survey (20 year olds in 2008)

Notes: All statistics are weighted. Parental migrant status is reported by the young person; all other parental characteristics are reported by the parent. *, ** and *** denote sample means that are significantly different from the column to the left at the 10%, 5% and 1% level respectively.

Table 2: Effects of suspension

	(1)	(2a)	(2b)	(3)	(4)
Report of schooling experiences	Own	Own	Own	Own	Mother's
(1) High school completion	-0.257***	-0.181***	-0.201***	-0.186***	-0.193***
	(0.020)	(0.022)	(0.032)	(0.031)	(0.037)
Observations	3,525	3,003	1,505	1,505	1,505
Pseudo R-squared	0.133	0.168	0.144	0.162	0.154
(2) Achieving a university entrance score	-0.139***	-0.098***	-0.091**	-0.085**	-0.132***
	(0.028)	(0.028)	(0.039)	(0.039)	(0.050)
Observations	2,514	2,214	1,161	1,161	1,162
Pseudo R-squared	0.049	0.081	0.090	0.106	0.108
(3) University entrance score	-7.638***	-6.833***	-6.807***	-7.209***	-6.756***
	(1.342)	(1.399)	(1.910)	(1.906)	(2.470)
Observations	1,818	1,635	891	891	892
R-squared	0.035	0.081	0.097	0.123	0.114
Other control variables					
Other schooling experiences	Yes	Yes	Yes	Yes	Yes
Own characteristics	Yes	Yes	Yes	Yes	Yes
Family welfare history	No	Yes	Yes	Yes	Yes
Family characteristics at age 14	No	Yes	Yes	Yes	Yes
Parental characteristics	No	No	No	Yes	Yes

Notes: For panels 1-2, average marginal effects are calculated from estimated probit coefficients. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Effect of suspension on outcome under varying degrees of correlation between the unobserved determinants of the two

	ρ =0	ρ =-0.1	ρ =-0.2	ρ =-0.3	ρ =-0.4	ρ =-0.5
(1) Own report of schooling experiences						
High school completion	-0.652***	-0.475***	-0.296***	-0.116*	0.065	0.247***
	(0.069)	(0.068)	(0.068)	(0.066)	(0.065)	(0.062)
	[-0.181]	[-0.127]	[-0.076]	[-0.029]	[0.015]	[0.057]
Achieving a university entrance score	-0.353***	-0.169*	0.015	0.197**	0.378***	0.556***
	(0.094)	(0.093)	(0.092)	(0.090)	(0.088)	(0.085)
	[-0.098]	[-0.044]	[0.004]	[0.046]	[0.084]	[0.118]
University entrance score	-6.833***	-3.904***	-0.939	2.108	5.294***	8.703***
	(1.392)	(1.390)	(1.386)	(1.379)	(1.369)	(1.355)
(2) Mother's report of schooling experiences						
High school completion	-0.714***	-0.527***	-0.338***	-0.149	0.042	0.233**
	(0.117)	(0.116)	(0.115)	(0.113)	(0.110)	(0.106)
	[-0.193]	[-0.135]	[-0.082]	[-0.034]	[0.009]	[0.048]
Achieving a university entrance score	-0.489***	-0.295*	-0.102	0.091	0.283*	0.473***
	(0.164)	(0.163)	(0.161)	(0.158)	(0.153)	(0.148)
	[-0.132]	[-0.075]	[-0.024]	[0.020]	[0.059]	[0.092]
University entrance score	-6.756***	-3.746	-0.705	2.402	5.625**	9.036***
	(2.434)	(2.429)	(2.416)	(2.393)	(2.360)	(2.315)

Notes: Standard errors are in parentheses, marginal effects in square brackets. *** p<0.01, ** p<0.05, * p<0.1. Shaded cells show the values of the correlation coefficient at which the effect of suspension loses statistical significance.

Table 4: Amount of selection on unobservables relative to selection on observables required to attribute the entire effect of suspension to selection bias

	High school completion	Achieving a university entrance score	University entrance score
(1) Own report of schooling experiences			
Unconstrained estimate of suspension	-0.652***	-0.353***	-6.833***
Standard error	(0.069)	(0.094)	(1.399)
Marginal effect	[-0.181]	[-0.098]	
Implied bias	-1.346	-1.223	-23.777
Ratio of estimate to bias	0.484	0.289	0.287
Sample size	3,003	2,214	1,635
Number of suspended respondents	529	259	150
(2) Mother's report of schooling experiences			
Unconstrained estimate of suspension	-0.714***	-0.489***	-6.756***
Standard error	(0.117)	(0.164)	(2.470)
Marginal effect	[-0.193]	[-0.132]	
Implied bias	-1.677	-1.264	-16.661
Ratio of estimate to bias	0.426	0.387	0.405
Sample size	1,505	1,162	892
Number of suspended respondents	161	79	44

Notes: Standard errors are in parentheses, marginal effects in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Effects of suspension, by socio-economic circumstances

	High school completion	Achieving a university entrance score	University entrance score
	(1)	(2)	(3)
(1) Own report of schooling experiences			
(a) If socio-economic disadvantage = intensive family welfare receipt			
For non-disadvantaged (1)	-0.166***	-0.108***	-7.153***
	(0.027)	(0.034)	(1.662)
For disadvantaged (2)	-0.212***	-0.087*	-5.893**
	(0.034)	(0.050)	(2.566)
(2) – (1) (interaction effect)	-0.046	0.022	1.26
	(0.043)	(0.060)	(3.049)
Observations	3003	2214	2214
(Pseudo) R-squared	0.165	0.077	0.08
Number of disadvantaged suspended	231	88	46
(b) If socio-economic disadvantage = any family welfare receipt			
For non-disadvantaged (1)	-0.165***	-0.038	-5.393**
	(0.043)	(0.045)	(2.231)
For disadvantaged (2)	-0.194***	-0.130***	-7.798***
	(0.025)	(0.036)	(1.773)
(2) – (1) (interaction effect)	-0.029	-0.092	-2.405
	(0.049)	(0.057)	(2.836)
Observations	3003	2214	2214
(Pseudo) R-squared	0.168	0.08	0.081
Number of disadvantaged suspended	417	184	93
(2) Mother's report of schooling experiences			
(a) If socio-economic disadvantage = intensive family welfare receipt			
For non-disadvantaged (1)	-0.253***	-0.137**	-5.143*
	(0.048)	(0.061)	(2.889)
For disadvantaged (2)	-0.141**	-0.132	-12.266***
	(0.057)	(0.089)	(4.521)
(2) – (1) (interaction effect)	0.112	0.005	-7.123
	(0.073)	(0.107)	(5.363)
Observations	1573	1207	1207
(Pseudo) R-squared	0.143	0.078	0.093
Number of disadvantaged suspended	69	31	14
(b) If socio-economic disadvantage = any family welfare receipt			
For non-disadvantaged (1)	-0.226***	-0.137	-3.896
	(0.080)	(0.094)	(4.195)
For disadvantaged (2)	-0.205***	-0.139**	-8.814***
	(0.042)	(0.060)	(2.980)
(2) – (1) (interaction effect)	0.022	-0.002	-4.918
	(0.090)	(0.111)	(5.136)
Observations	1573	1207	1207
(Pseudo) R-squared	0.145	0.08	0.093
Number of disadvantaged suspended	136	63	31

Notes: For columns 1-2, average marginal effects are calculated from estimated probit coefficients. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All regressions control for other schooling experiences, own characteristics, family characteristics at age 14, and parental characteristics as in column (3) of Table 2.

Appendix Table 1: Effects on the probability of completing high school

	(1)	(2a)	(2b)	(3)	(4)
Report of schooling experiences	Own	Own	Own	Own	Mother's
Suspended from school	-0.257***	-0.181***	-0.201***	-0.186***	-0.193***
	(0.020)	(0.022)	(0.032)	(0.031)	(0.037)
Other schooling experiences					
Repeated a year	-0.168***	-0.143***	-0.165***	-0.149***	-0.152***
	(0.024)	(0.025)	(0.039)	(0.037)	(0.038)
Number of schools	-0.026***	-0.011**	-0.004	-0.003	-0.009
	(0.005)	(0.005)	(0.008)	(0.008)	(0.008)
Own characteristics					
Female	0.025*	0.043***	0.037**	0.035*	0.040**
	(0.013)	(0.014)	(0.019)	(0.018)	(0.018)
Indigenous Australian	-0.167***	-0.126***	-0.132**	-0.090	-0.089
	(0.040)	(0.040)	(0.062)	(0.058)	(0.057)
Born in a NES country	0.115***	0.086***	0.090**	0.061	0.060
	(0.020)	(0.022)	(0.038)	(0.050)	(0.050)
Metropolitan residence	0.078***	0.064***	0.046**	0.031	0.033*
	(0.014)	(0.015)	(0.019)	(0.019)	(0.019)
Family characteristics at age 14					
Lived with both parents at 14		0.060***	-0.009	-0.016	-0.013
		(0.018)	(0.024)	(0.024)	(0.024)
Mother employed at 14		0.024*	0.032	0.022	0.024
		(0.015)	(0.021)	(0.021)	(0.021)
Mother had Year 12 when youth 14		0.051***	0.033*	0.011	0.012
		(0.015)	(0.019)	(0.021)	(0.021)
Mother had Yr 12 when youth 14: missing		-0.077**	-0.093*	-0.086*	-0.083*
		(0.032)	(0.053)	(0.051)	(0.050)
Father had Year 12 when youth 14		0.045***	0.025	0.006	0.009
		(0.015)	(0.020)	(0.022)	(0.022)
Father had Yr 12 when youth 14: missing		-0.082***	-0.062	-0.061	-0.068
		(0.031)	(0.046)	(0.044)	(0.045)
Family welfare history					
Moderate (late) receipt		-0.059***	-0.063**	-0.056*	-0.052*
		(0.023)	(0.031)	(0.031)	(0.031)
Moderate (early) receipt		-0.052***	-0.050**	-0.038	-0.032
		(0.018)	(0.023)	(0.023)	(0.023)
Intensive receipt		-0.078***	-0.073***	-0.053**	-0.049*
		(0.019)	(0.027)	(0.027)	(0.027)
Parental characteristics					
Age of mother				0.003	0.002
				(0.002)	(0.002)
Number of children of mother				-0.014**	-0.011
				(0.007)	(0.007)
Mother is a smoker				-0.024	-0.041*
				(0.024)	(0.024)
At least one parent born in a NES country				0.037	0.038
				(0.025)	(0.025)
At least one parent born in NES: missing				-0.058	-0.046
				(0.148)	(0.141)
Mother's education: Year 12				0.058*	0.054
				(0.034)	(0.034)
Mother's education: above Year 12				0.017	0.019
				(0.022)	(0.023)
Father's education: Year 12				0.026	0.037
				(0.031)	(0.031)
Father's education: above Year 12				0.063***	0.065***
				(0.023)	(0.023)
Observations	3525	3003	1505	1505	1505
Pseudo R-squared	0.133	0.168	0.144	0.162	0.154

Notes: Average marginal effects are calculated from estimated probit coefficients. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 2: Effects on the probability of achieving a university entrance score, the university entrance score achieved and the probability of suspension

	Achieving a university entrance score		University entrance score		Suspension
	(1)	(2)	(3)	(4)	(5)
Report of schooling experiences	Own	Mother's	Own	Mother's	Own
Suspended from school	-0.098***	-0.132***	-6.833***	-6.756***	
	(0.028)	(0.050)	(1.399)	(2.470)	
Other schooling experiences					
Repeated a year	-0.187***	-0.092*	-5.758***	-3.267	0.062*
	(0.039)	(0.048)	(1.928)	(2.626)	(0.034)
Number of schools	-0.003	-0.014	-0.004	-0.588	0.031***
	(0.007)	(0.010)	(0.364)	(0.519)	(0.008)
Own characteristics					
Female	0.056***	0.046**	2.760***	4.058***	-0.106***
	(0.017)	(0.022)	(0.818)	(1.083)	(0.019)
Indigenous Australian	-0.022	0.005	-5.317	-0.630	0.051
	(0.053)	(0.066)	(3.485)	(4.298)	(0.056)
Born in a NES country	0.064**	0.046	1.038	2.071	-0.085**
	(0.026)	(0.054)	(1.396)	(2.892)	(0.040)
Metropolitan residence	0.067***	0.035	1.815**	2.431**	0.011
	(0.019)	(0.023)	(0.916)	(1.164)	(0.019)
Family characteristics at age 14					
Lived with both parents at 14	-0.002	-0.056**	0.608	1.233	-0.025
	(0.022)	(0.026)	(1.178)	(1.650)	(0.027)
Mother employed at 14	0.029	0.001	-0.352	-1.274	-0.011
	(0.018)	(0.025)	(0.925)	(1.303)	(0.021)
Mother had Year 12 when youth 14	0.032*	0.001	3.961***	1.836	-0.020
	(0.018)	(0.026)	(0.890)	(1.303)	(0.022)
Mother had Yr 12 when youth 14: missing	-0.085*	-0.079	0.182	-1.339	0.003
	(0.046)	(0.067)	(2.374)	(3.735)	(0.045)
Father had Year 12 when youth 14	0.046**	0.047*	4.218***	3.378***	0.008
	(0.018)	(0.026)	(0.895)	(1.303)	(0.022)
Father had Yr 12 when youth 14: missing	-0.080*	-0.084	2.106	0.765	0.054
	(0.045)	(0.066)	(2.284)	(3.210)	(0.044)
Family welfare history					
Moderate (late) receipt	-0.025	0.002	-1.609	-0.876	0.026
	(0.025)	(0.031)	(1.251)	(1.648)	(0.031)
Moderate (early) receipt	-0.061***	-0.082***	-1.345	-0.258	0.018
	(0.021)	(0.027)	(1.047)	(1.381)	(0.023)
Intensive receipt	-0.087***	-0.109***	-1.778	0.609	0.060**
	(0.024)	(0.035)	(1.181)	(1.710)	(0.028)
Parental characteristics					
Age of mother		0.004*		0.301**	0.003
		(0.002)		(0.119)	(0.002)
Number of children of mother		-0.006		-0.357	0.003
		(0.009)		(0.458)	(0.007)
Mother is a smoker		-0.075**		-1.611	0.095***
		(0.031)		(1.617)	(0.027)
At least one parent born in NES		0.013		1.820	-0.005
		(0.030)		(1.536)	(0.026)
At least one parent born in NES: missing		-0.104		-16.364	-0.048
		(0.292)		(16.186)	(0.104)
Mother's education: Year 12		0.024		2.294	-0.022
		(0.044)		(2.263)	(0.033)
Mother's education: above Year 12		0.026		0.969	0.039*
		(0.029)		(1.474)	(0.022)
Father's education: Year 12		-0.005		-1.883	-0.103***
		(0.036)		(1.811)	(0.029)
Father's education: above Year 12		0.030		3.069**	-0.096***
		(0.027)		(1.402)	(0.024)
Observations	2214	1162	1635	892	1505
(Pseudo) R-squared	0.081	0.108	0.081	0.114	0.113

Notes: For columns 1, 2 and 5, average marginal effects are calculated from estimated probit coefficients. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.